Sorting out the Most Confusing English Phrasal Verbs

Yuancheng Tu
Department of Linguistics
University of Illinois
ytu@illinois.edu

Dan Roth
Department of Computer Science
University of Illinois
danr@illinois.edu

Abstract

In this paper, we investigate a full-fledged supervised machine learning framework for identifying English phrasal verbs in a given context. We concentrate on those that we define as *the most confusing* phrasal verbs, in the sense that they are the most commonly used ones whose occurrence may correspond either to a true phrasal verb or an alignment of a simple verb with a preposition.

We construct a benchmark dataset[1] with 1,348 sentences from BNC, annotated via an Internet crowdsourcing platform. This dataset is further split into two groups, more *idiomatic* group which consists of those that tend to be used as a true phrasal verb and more *compositional* group which tends to be used either way. We build a discriminative classifier with easily available lexical and syntactic features and test it over the datasets. The classifier overall achieves 79.4% accuracy, 41.1% error deduction compared to the corpus majority baseline 65%. However, it is even more interesting to discover that the classifier learns more from the more *compositional* examples than those *idiomatic* ones.

1 Introduction

Phrasal verbs in English (also called English Particle Constructions), are syntactically defined as combinations of verbs and prepositions or particles, but semantically their meanings are generally not the direct sum of their parts. For example, *give in* means *submit, yield* in the sentence, *Adam's saying it's important to stand firm , not give in to terrorists.* Adam was not *giving* anything and he was not *in* anywhere either. Kolln and Funk (Kolln and Funk, 1998) use *the test of meaning* to detect English phrasal verbs, i.e., each phrasal verb could be replaced by a single verb with the same general meaning, for example, using *yield* to replace *give in* in the aforementioned sentence. To confuse the issue even further, some phrasal verbs, for example, *give in* in the following two sentences, are used either as a true phrasal verb (the first sentence) or not (the second sentence) though their surface forms look cosmetically identical.

1. How many Englishmen **gave in** to their emotions like that ?
2. It is just this denial of anything beyond what is directly **given in** experience that marks Berkeley out as an empiricist .

This paper is targeting to build an automatic learner which can recognize a true phrasal verb from its orthographically identical construction with a verb and a prepositional phrase. Similar to other types of MultiWord Expressions (MWEs) (Sag et al., 2002), the syntactic complexity and semantic idiosyncrasies of phrasal verbs pose many particular challenges in empirical Natural Language Processing (NLP). Even though a few of previous works have explored this identification problem empirically (Li et al., 2003; Kim and Baldwin, 2009) and theoretically (Jackendoff, 2002), we argue in this paper that this context sensitive identification problem is not so easy as conceivably shown before, especially when it is used to handle those more *compositional* phrasal verbs which are empirically used either way in the corpus as a true phrasal verb or a simplex verb with a preposition combination. In

[1] http://cogcomp.cs.illinois.edu/page/resources/PVC_Data

addition, there is still a lack of adequate resources or benchmark datasets to identify and treat phrasal verbs within a given context. This research is also an attempt to bridge this gap by constructing a publicly available dataset which focuses on some of the most commonly used phrasal verbs within their most confusing contexts.

Our study in this paper focuses on six of the most frequently used verbs, *take*, *make*, *have*, *get*, *do* and *give* and their combination with nineteen common prepositions or particles, such as *on*, *in*, *up*, *for*, *over* etc. We categorize these phrasal verbs according to their continuum of compositionality, splitting them into two groups based on the biggest gap within this scale, and build a discriminative learner which uses easily available syntactic and lexical features to analyze them comparatively. This learner achieves 79.4% overall accuracy for the whole dataset and learns the most from the more *compositional* data group with 51.2% error reduction over its 46.6% majority baseline.

2 Related Work

Phrasal verbs in English were observed as one kind of composition that is used frequently and constitutes the greatest difficulty for language learners more than two hundred and fifty years ago in Samuel Johnson's *Dictionary of English Language*[2]. They have also been well-studied in modern linguistics since early days (Bolinger, 1971; Kolln and Funk, 1998; Jackendoff, 2002). Careful linguistic descriptions and investigations reveal a wide range of English phrasal verbs that are syntactically uniform, but diverge largely in semantics, argument structure and lexical status. The complexity and idiosyncrasies of English phrasal verbs also pose a special challenge to computational linguistics and attract considerable amount of interest and investigation for their extraction, disambiguation as well as identification.

Recent computational research on English phrasal verbs have been focused on increasing the coverage and scalability of phrasal verbs by either extracting unlisted phrasal verbs from large corpora (Villavicencio, 2003; Villavicencio, 2006), or constructing productive lexical rules to generate new cases (Villanvicencio and Copestake, 2003). Some other researchers follow the semantic regularities of the particles associated with these phrasal verbs and concentrate on disambiguation of phrasal verb semantics, such as the investigation of the most common particle *up* by (Cook and Stevenson, 2006).

Research on token identification of phrasal verbs is much less compared to the extraction. (Li et al., 2003) describes a regular expression based simple system. Regular expression based method requires human constructed regular patterns and cannot make predictions for *Out-Of-Vocabulary* phrasal verbs. Thus, it is hard to be adapted to other NLP applications directly. (Kim and Baldwin, 2009) proposes a memory-based system with post-processed linguistic features such as selectional preferences. Their system assumes the perfect outputs of a parser and requires laborious human corrections to them.

The research presented in this paper differs from these previous identification works mainly in two aspects. First of all, our learning system is fully automatic in the sense that no human intervention is needed, no need to construct regular patterns or to correct parser mistakes. Secondly, we focus our attention on the comparison of the two groups of phrasal verbs, the more *idiomatic* group and the more *compositional* group. We argue that while more *idiomatic* phrasal verbs may be easier to identify and can have above 90% accuracy, there is still much room to learn for those more *compostional* phrasal verbs which tend to be used either positively or negatively depending on the given context.

3 Identification of English Phrasal Verbs

We formulate the context sensitive English phrasal verb identification task as a supervised binary classification problem. For each target candidate within a sentence, the classifier decides if it is a true phrasal verb or a simplex verb with a preposition. Formally, given a set of n labeled examples $\{x_i, y_i\}_{i=1}^{n}$, we learn a function $f : \mathcal{X} \rightarrow \mathcal{Y}$ where $\mathcal{Y} \in \{-1, 1\}$. The learning algorithm we use is the soft-margin SVM with L2-loss. The learning package we use is LIBLINEAR (Chang and Lin, 2001)[3].

Three types of features are used in this discriminative model. (1)*Words*: given the window size from

[2] It is written in the Preface of that dictionary.

[3] http://www.csie.ntu.edu.tw/∼cjlin/liblinear/

the one before to the one after the target phrase, *Words* feature consists of every surface string of all shallow chunks within that window. It can be an n-word chunk or a single word depending on the the chunk's bracketing. (2)*ChunkLabel*: the chunk name with the given window size, such as *VP*, *PP*, etc. (3)*ParserBigram*: the bi-gram of the non-terminal label of the parents of both the verb and the particle. For example, from this partial tree *(VP (VB get)(PP (IN through)(NP (DT the)(NN day)))*, the parent label for the verb *get* is *VP* and the parent node label for the particle *through* is *PP*. Thus, this feature value is *VP-PP*. Our feature extractor is implemented in Java through a publicly available NLP library[4] via the tool called Curator (Clarke et al., 2012). The shallow parser is publicly available (Punyakanok and Roth, 2001)[5] and the parser we use is from Charniak (Charniak and Johnson, 2005).

4 Experiments and Analysis

In this section, we first present the generation and annotation of our phrasal verb dataset and the criteria we use to split this dataset into two groups: more *idiomatic* group and more *compositional* group. Then we describe in detail our experiment results with respect to both of whole dataset as well as these two groups.

4.1 Data Preparation and Annotation

All sentences in our dataset are extracted from BNC (XML Edition), a balanced synchronic corpus containing 100 million words collected from various sources of British English. We first construct a list of phrasal verbs for the six verbs that we are interested in from two resources, WN3.0 (Fellbaum, 1998) and DIRECT[6]. Since these targeted verbs are also commonly used in English Light Verb Constructions (LVCs), we filter out LVCs in our list using a publicly available LVC corpus (Tu and Roth, 2011). The result list consists of a total of 245 phrasal verbs. We then search over BNC and find sentences for all of them. We choose the frequency threshold to be 25 and generate a list of 122 phrasal verbs. Finally we manually pick out 23 of these phrasal verbs and sample randomly 10% extracted sentences for each of them for annotation.

The annotation is done through a crowdsourcing platform[7]. The annotators are asked to identify true phrasal verbs within a sentence. The reported inner-annotator agreement is 84.5% and the gold average accuracy is 88%. These numbers indicate the good quality of the annotation. The final corpus consists of 1,348 sentences among which, 65% with a true phrasal verb and 35% with a simplex verb-preposition combination.

4.2 Dataset Splitting

Table 1 lists all verbs in the dataset. *Total* is the total number of sentences annotated for that phrasal verb and *Positive* indicated the number of examples which are annotated as containing the true phrasal verb usage. In this table, the decreasing percentage of the true phrasal verb usage within the dataset indicates the increasing compositionality of these phrasal verbs. The natural division line with this scale is the biggest percentage gap (about 10%) between *make_out* and *get_at*. Hence, two groups are split over that gap. The more *idiomatic* group consists of the first 11 verbs with 554 sentences and 91% of these sentences include true phrasal verb usage. This data group is more biased toward the positive examples. The more *compositional* data group has 12 verbs with 794 examples and only 46.6% of them contain true phrasal verb usage. Therefore, this data group is more balanced with respective to positive and negative usage of the phrase verbs.

4.3 Experimental Results and Discussion

Our results are computed via 5-cross validation. We plot the classifier performance with respect to the overall dataset, the more *compositional* group and the more *idiomatic* group in Figure 1. The classifier only improves 0.6% when evaluated on the *idiomatic* group. Phrasal verbs in this dataset are more biased toward behaving like an idiom regardless of their contexts, thus are more likely to be captured by rules or patterns. We assume this may explain some high numbers reported in some previous works. However, our classifier is more effective over the more *compositional* group and reaches 73.9% accuracy, a 51.1% error deduction comparing to its majority baseline. Phrasal verbs in this set tend to be used

[4] http://cogcomp.cs.illinois.edu/software/edison/

[5] http://cogcomp.cs.illinois.edu/page/software_view/Chunker

[6] http://u.cs.biu.ac.il/∼nlp/downloads/DIRECT.html

[7] crowdflower.com

Verb	Total	Positive	Percent(%)
get_onto	6	6	1.00
get_through	61	60	0.98
get_together	28	27	0.96
get_on_with	70	67	0.96
get_down_to	17	16	0.94
get_by	11	10	0.91
get_off	51	45	0.88
get_behind	7	6	0.86
take_on	212	181	0.85
get_over	34	29	0.85
make_out	57	48	0.84
get_at	35	26	0.74
get_on	142	103	0.73
take_after	10	7	0.70
do_up	13	8	0.62
get_out	206	118	0.57
do_good	8	4	0.50
make_for	140	65	0.46
get_it_on	9	3	0.33
get_about	20	6	0.30
make_over	12	3	0.25
give_in	118	27	0.23
have_on	81	13	0.16
Total: 23	1348	878	0.65

Table 1: The top group consists of the more *idiomatic* phrasal verbs with 91% of their occurrence within the dataset to be a true phrasal verb. The second group consists of those more *compositional* ones with only 46.6% of their usage in the dataset to be a true phrasal verb.

equally likely as a true phrasal verb and as a simplex verb-preposition combination, depending on their context. We argue phrasal verbs such as these pose a real challenge for building an automatic context sensitive phrasal verb classifier. The overall accuracy of our preliminary classifier is about 79.4% when it is evaluated over all examples from these two groups.

Finally, we conduct an ablation analysis to explore the contributions of the three types of features in our model and their accuracies with respect to each data group are listed in Table 2 with the boldfaced best performance. Each type of features is used individually in the classifier. The feature type *Words* is the most effective feature with respect to the *idiomatic* group and the overall dataset. And the *chunk* feature is more effective towards the *compositional* group, which may explain the linguistic intuition that negative phrasal verbs usually do not belong to the same syntactic chunk.

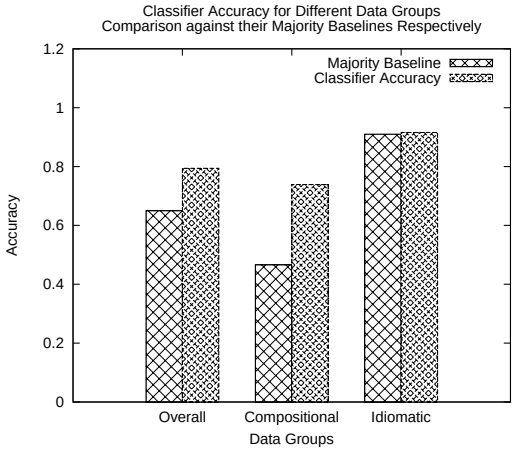

Figure 1: Classifier Accuracy of each data group, comparing with their baseline respectively. Classifier learns the most from the more *compositional* group, indicated by its biggest histogram gap.

	Datasets		
	Overall	Compositional	Idiom.
Baseline	65.0%	46.6%	91%
Words	**78.6**%	70.2%	**91.4**%
Chunk	65.6%	**70.7**%	89.4%
ParserBi	64.4%	67.2%	89.4%

Table 2: Accuracies achieved by the classifier when tested on different data groups. Features are used individually to evaluate the effectiveness of each type.

5 Conclusion

In this paper, we build a discriminative learner to identify English phrasal verbs within a given context. By focusing our attention on the comparison between these more idiomatically biased phrasal verbs and those more *compositional* ones, we are able to not only explain the conceivable high accuracy a classifier may achieve over these more *idiomatic* ones, but also arguing that the bigger challenge is for those more *compositional* cases.

Our contributions in this paper are threefold. We construct a publicly available context sensitive English phrasal verb dataset with 1,348 sentences from BNC. We split the dataset into two groups according to their tendency toward idiosyncrasy and compositionality, and build a discriminative learner which uses easily available syntactic and lexical features to analyze them comparatively. We demonstrate empirically that high accuracy achieved by models may be due to the stronger idiomatic tendency of these phrasal verbs. For many of the more *ambiguous* cases, a classifier learns more from the *compositional* examples and these phrasal verbs are more challenging.

Acknowledgments

The authors would like to thank four annonymous reviewers for their valuable comments. The research in this paper was supported by the Multimodal Information Access & Synthesis Center at UIUC, part of CCICADA, a DHS Science and Technology Center of Excellence and the Defense Advanced Research Projects Agency (DARPA) Machine Reading Program under Air Force Research Laboratory (AFRL) prime contract no. FA8750-09-C-0181. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the view of DHS, DARPA, AFRL, or the US government.

References

D. Bolinger. 1971. *The Phrasal Verb in English*. Harvard University Press.

C. Chang and C. Lin, 2001. *LIBSVM: a library for support vector machines*. Software available at http://www.csie.ntu.edu.tw/∼cjlin/libsvm.

E. Charniak and M. Johnson. 2005. Coarse-to-fine n-best parsing and maxent discriminative reranking. In *Proceedings of ACL-2005*.

J. Clarke, V. Srikumar, M. Sammons, and D. Roth. 2012. An NLP curator: How I learned to stop worrying and love NLP pipelines. In *Proceedings of LREC-2012*.

P. Cook and S. Stevenson. 2006. Classifying particle semantics in English verb-particle constructions. In *Proceedings of the Workshop on Multiword Expressions: Identifying and Exploiting Underlying Properties*, pages 45–53, Sydney, Australia.

C. Fellbaum, editor. 1998. *WordNet: An Electronic Lexical Database*. MIT Press.

R. Jackendoff. 2002. English particle constructions, the lexicon, and the autonomy of syntax. In N. Dehé, R. Jackendoff, A. McIntyre, and S. Urban, editors, *Verb-Particle Explorations*, pages 67–94. Mouton de Gruyter.

S Kim and T. Baldwin. 2009. How to pick out token instances of English verb-particle constructions. *Journal of Language Resources and Evaluation*.

M. Kolln and R. Funk. 1998. *Understanding English Grammar*. Allyn and Bacon.

W. Li, X. Zhang, C. Niu, Y. Jiang, and R. Srihari. 2003. An expert lexicon approach to identifying English phrasal verbs. In *Proceedings of the 41st Annual Meeting of ACL*, pages 513–520.

V. Punyakanok and D. Roth. 2001. The use of classifiers in sequential inference. In *NIPS*, pages 995–1001.

I. Sag, T. Baldwin, F. Bond, and A. Copestake. 2002. Multiword expressions: A pain in the neck for NLP. In *Proc. of the 3rd International Conference on Intelligent Text Processing and Computational Linguistics (CICLing-2002)*, pages 1–15.

Y. Tu and D. Roth. 2011. Learning english light verb constructions: Contextual or statistica. In *Proceedings of the ACL Workshop on Multiword Expressions: from Parsing and Generation to the Real World*.

A. Villanvicencio and A. Copestake. 2003. Verb-particle constructions in a computational grammar of English. In *Proceedings of the 9th International Conference on HPSG*, pages 357–371.

A. Villavicencio. 2003. Verb-particle constructions and lexical resources. In *Proceedings of the ACL 2003 Workshop on Multiword Expressions: Analysis, Acquisition and Treatment*, pages 57–64.

A. Villavicencio, 2006. *Computational Linguistics Dimensions of the Syntax and Semantics of Prepositions*, chapter Verb-Particel Constructions in the World Wide Web. Springer.